

**НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ
імені ІГОРЯ СІКОРСЬКОГО»**

**Інститут телекомунікаційних систем
Кафедра Інформаційно-телекомунікаційних мереж**

До захисту допущено:

Завідувач кафедри

_____ Лариса ГЛОБА

«__» _____ 2020 р.

Дипломна робота

на здобуття ступеня бакалавра

**за освітньо-професійною програмою «Інформаційно-комунікаційні
технології»**

спеціальності 172 «Телекомунікації та радіотехніка»

**на тему: «Підхід до дослідження психофізіологічного стану людини на
основі кольорової преференції та суб'єктивних відчуттів»**

Виконав :

студент IV курсу, групи ТІ-62

Копа Роман Вікторович _____

Керівник:

доцент кафедри ІТМ ІТС, к.т.н.,

Штогріна Олена Сергіївна _____

Рецензент:

професор кафедри ТК ІТС, доцент, к.т.н.

Якорнов Євгеній Аркадійович _____

Засвідчую, що у цій дипломній роботі
немає запозичень з праць інших авторів
без відповідних посилань.

Студент _____

Київ – 2020 року

Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»
Інститут телекомунікаційних систем
Кафедра Інформаційно-телекомунікаційних мереж

Рівень вищої освіти – перший (бакалаврський)

Спеціальність – 172 «Телекомунікації та радіотехніка»

Освітньо-професійна програма «Інформаційно-комунікаційні технології»

ЗАТВЕРДЖУЮ

Завідувач кафедри

_____ Лариса ГЛОБА

«__» _____ 2020 р.

ЗАВДАННЯ
на дипломну роботу студенту

1. Тема роботи «Підхід до дослідження психофізіологічного стану людини на основі кольорової преференції та суб'єктивних відчуттів», керівник роботи Штогріна Олена Сергіївна, доцент кафедри інформаційно-телекомунікаційних мереж, к.т.н., затверджені наказом по університету від «30» березня 2020 р. № 924-с

2. Термін подання студентом роботи 8 червня 2020 р.

3. Вихідні дані до роботи: результати проходження тестів кольорової преференцій «Біоколор 1» та «Біоколор 2», оцінка суб'єктивного стану зимівником, еталонне значення кольорової преференції.

4. Зміст роботи:

4.1. Провести огляд дослідження психофізіологічного стану людини на основі кольорової преференції та аналіз сучасних технологій обробки даних.

4.2. Запропонувати підхід до аналізу психофізіологічного стану людини на основі кольорової преференції та суб'єктивних відчуттів людини.

4.3. Дослідити ефективність методів машинного навчання для прогнозування стану людини, який базується на суб'єктивних відчуттях, на основі результатів тесту кольорової преференції та еталонного значення кольорової преференції.

4.4. Реалізувати програмне забезпечення запропонованого підходу в вигляді окремого модуля, для можливості його інтеграції в стороннє ПЗ.

5. Перелік ілюстративного матеріалу (із зазначенням плакатів, презентацій тощо)

1. Тема, актуальність, мета, задачі.

2. Огляд дослідження психофізіологічного стану людини на основі кольорової преференції та аналіз сучасних технологій обробки даних.

3. Підхід до аналізу психофізіологічного стану людини.

4. Дослідження ефективності методів машинного навчання для прогнозування стану людини.

5. Схема підходу до аналізу психофізіологічного стану людини на основі кольорової преференції. Результат роботи.

6. Загальні висновки.

6. Дата видачі завдання 10 вересня 2019 р.

Календарний план

№ з/п	Назва етапів виконання дипломної роботи	Термін виконання етапів роботи	Примітка
1	Проведення огляду дослідження психофізіологічного стану людини.	10.09.2019-10.10.2019	виконано
2	Аналіз сучасних технологій обробки даних.	11.10.2019-11.11.2019	виконано
3	Розробка підходу до аналізу психофізіологічного стану людини.	12.11.2019-12.01.2020	виконано
4	Дослідження ефективності методів машинного навчання для прогнозування стану людини.	13.01.2020-21.03.2020	виконано

5	Реалізація програмного забезпечення.	21.03.2020-01.05.2020	виконано
---	--------------------------------------	-----------------------	----------

Студент

Роман КОПА

Керівник роботи

Олена ШТОГРІНА

РЕФЕРАТ

Робота містить 58 сторінки, 25 рисунків та 1 таблицю. Було використано 6 джерел.

Мета роботи: підвищити швидкість та ефективність прогнозування психофізіологічного стану людини за рахунок розробки підходу до аналізу психофізіологічного стану людини на основі кольорової преференції та суб'єктивних відчуттів людини.

Розглянуто методи дослідження психофізіологічного стану людини на основі кольорові преференції та проведено аналіз задач машинного навчання для обробки та аналізу даних. Описано проблеми дослідження психофізіологічного стану людини, визначено необхідність у розробці підходу до прогнозування стану людини. Досліджено ефективність різних методів машинного навчання для прогнозування стану людини. Запропоновано підхід до аналізу психофізіологічного стану людини на основі тесту кольорової преференції та суб'єктивних відчуттів людини. Застосування запропонованого підходу дозволяє, подавши на вхід послідовність тесту преференцій, еталонне значення результуючої преференції отримати на виході оцінку стану людини, яка базується на суб'єктивному стані людини. Це дозволяє завчасно виявляти та реагувати на погіршення психофізіологічного стану людини.

Ключові слова: кольорова преференція, тест преференцій, прогнозування стану, психофізіологічний стан, зимівник.

ABSTRACT

The work contains 58 pages, 25 figures and 1 table, 6 sources have been used.

Goal: to increase the speed and efficiency of forecasting the psychophysiological state of a person by developing an approach to the analysis of the psychophysiological state of a person based on color preference and subjective feelings of a person.

Methods of research of a psychophysiological condition of the person on the basis of color preferences are considered and the analysis of problems of machine learning for data processing and analysis is carried out. The problems of research of a psychophysiological condition of the person are described, necessity in development of the approach to forecasting of a condition of the person is defined. The effectiveness of various methods of machine learning for predicting the human condition has been studied. An approach to the analysis of the psychophysiological state of a person on the basis of the test of color preference and the reference value of color preference is proposed. The application of the proposed approach allows, by entering the sequence of the test of preferences, the reference value of the resulting preference to obtain an estimate of the human condition, which is based on the subjective state of man. This allows you to detect and respond in advance to the deterioration of the psychophysiological state of man.

Key words: color preference, test of preferences, prognosis, psychophysiological state, winter time.

ЗМІСТ

ВСТУП.....	10
РОЗДІЛ 1.....	12
ОГЛЯД ДОСЛІДЖЕННЯ ПСИХОФІЗІОЛОГІЧНОГО СТАНУ ЛЮДИНИ НА ОСНОВІ КОЛЬОРОВОЇ ПРЕФЕРЕНЦІЇ ТА АНАЛІЗ СУЧАСНИХ ТЕХНОЛОГІЙ ОБРОБКИ ДАНИХ.....	12
1.1. Метод дослідження психофізіологічного стану на основі кольорової преференції.....	13
1.2. Аналіз сучасних технологій обробки даних	17
1.2.1. Задача класифікації.....	17
1.2.2. Задача регресії.....	19
1.2.3. Задача кластеризації	20
Висновки:.....	21
РОЗДІЛ 2.....	23
ПІДХІД ДО АНАЛІЗУ ПСИХОФІЗІОЛОГІЧНОГО СТАНУ ЛЮДИНИ НА ОСНОВІ КОЛЬОРОВОЇ ПРЕФЕРЕНЦІЇ	23
2.1. Аналіз психофізіологічного стану людини на основі кольорової преференції.....	23
2.2. Еталонний результат кольорової преференції	25
2.3. Підхід до аналізу психофізіологічного стану людини на основі кольорової преференції та суб'єктивного відчуття людини.....	26
ДОСЛІДЖЕННЯ ЕФЕКТИВНОСТІ МЕТОДІВ МАШИННОГО НАВЧАННЯ ДЛЯ ПРОГНОЗУВАННЯ СТАНУ ЛЮДИНИ.	30
3.1. Лінійна регресія	30

3.2. Регресія опорних векторів.....	33
3.3. Метод k-найближчих сусідів	39
3.4. Регресійне дерево	43
3.3. Порівняння ефективності методів машинного навчання для прогнозування стану людини.	49
Висновки:.....	51
РОЗДІЛ 4.....	52
РЕАЛІЗАЦІЯ ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ ЗАПРОПОНОВАНОГО ПІДХОДУ ДО АНАЛІЗУ ПСИХОФІЗІОЛОГІЧНОГО СТАНУ ЛЮДИНИ НА ОСНОВІ КОЛЬОРОВОЇ ПРЕФЕРЕНЦІЇ ТА СУБ'ЄКТИВНИХ ВІДЧУТІВ.....	52
4.1. Постановка завдання та вимоги до ПЗ.....	52
4.2. Опис розробленого ПЗ	52
4.3 Практичне застосування розробленого ПЗ	54
Висновки:.....	56
ЗАГАЛЬНІ ВИСНОВКИ ПО РОБОТІ.....	57
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ.....	58

ПЕРЕЛІК СКОРОЧЕНЬ

ПЗ

Програмне забезпечення

НАНЦ

Національний Антарктичний Науковий Центр

ВСТУП

Актуальність. Трудова діяльність в особливих умовах пов'язана з впливом на людину цілого комплексу екстремальних факторів, що супроводжується перенапругою адаптивних реакцій і розвитком стресових станів. Тому для роботи в екстремальних умовах пред'являються підвищені вимоги до психофізіологічного стану людини.

За таких екстремальних умов змушені працювати зимівники на антарктичній станції «Академік Вернадський», де протягом всього перебування зимівників фіксується велика кількість випадків порушень нормального режиму сну, симптомів головного болю, погіршення настрою і самопочуття. Дані випадки погіршення стану людини сприяють зростанню ризику порушення психофізіологічного стану зимівників.

Наразі, в НАНЦ України активно ведуться дослідження психофізіологічного стану людини методами на основі кольорової преференції. Методи на основі кольорової преференції являються громіздкими та потребують досвіду роботи з даними. Тому доцільно розробити підхід до аналізу психофізіологічного стану, який би спростив та пришвидшив роботу психологів.

Мета роботи: підвищити швидкість та ефективність прогнозування психофізіологічного стану людини за рахунок розробки підходу до аналізу психофізіологічного стану людини на основі кольорової преференції та суб'єктивних відчуттів людини.

Для досягнення мети дослідження було поставлено та вирішено такі основні задачі:

1. Провести огляд дослідження психофізіологічного стану людини на основі кольорової преференції та аналіз сучасних технологій обробки даних.

2. Запропонувати підхід до аналізу психофізіологічного стану людини на основі кольорової преференції та суб'єктивних відчуттів людини.
3. Дослідити ефективність методів машинного навчання для прогнозування стану людини, який базується на суб'єктивних відчуттях, на основі проходження тесту кольорової преференції та еталонного значення кольорової преференції.
4. Реалізувати програмне забезпечення запропонованого підходу в вигляді окремого модуля, для можливості його інтеграції в стороннє ПЗ.

Теоретичний результат дослідження:

1. Проведено огляд дослідження психофізіологічного стану людини на основі кольорової преференції та аналіз сучасних технологій обробки даних.
2. Запропоновано підхід до аналізу психофізіологічного стану людини на основі кольорової преференції та суб'єктивних відчуттів.

Практичний результат роботи:

1. Досліджено ефективність різних методів машинного навчання для прогнозування стану людини, який базується на суб'єктивних відчуттях.
2. Реалізовано програмне забезпечення запропонованого підходу.

РОЗДІЛ 1

ОГЛЯД ДОСЛІДЖЕННЯ ПСИХОФІЗІОЛОГІЧНОГО СТАНУ ЛЮДИНИ НА ОСНОВІ КОЛЬОРОВОЇ ПРЕФЕРЕНЦІЇ ТА АНАЛІЗ СУЧАСНИХ ТЕХНОЛОГІЙ ОБРОБКИ ДАНИХ.

Недолік світла, різноманітність кольорового оточення або його монохромність можуть негативно відбиватися на психоемоційному стані людини, викликати зміни в системі центральних механізмів регуляції з подальшим розвитком хронічного стресу, депресії і подальшим погіршенням психофізіологічного стану людини. З'ясування психічних і психофізіологічних особливостей людини особливо важливо для відбору фахівців, що працюють в екстремальних умовах, наприклад при відборі фахівців до експедиції на антарктичну станцію «Академік Вернадський».

Визначення особливостей психічних і психофізіологічних станів людини пов'язано з необхідністю враховувати безліч супутніх чинників, оскільки формування людини як особистості представляється динамічним процесом, який залежить від соціальних умов, середовища в якому перебуває людина, а також від адаптивних здібностей людини [1].

Багаторічні дослідження впливу спектрального ряду сонячного світла на психофізіологічні функції організму, аналіз відповідних реакцій відчуттів і сприйняття людини, а також вивчені історичні дані про вплив навколишнього колірної фону на еволюційні процеси розвитку суспільства були покладені в основу методики оцінювання поточного психофізіологічного статусу «БІОКОЛОР ». Це динамічна модель технології кольорового аналізу результатів дослідження кольорової преференції, що дозволяє оцінити поточний психофізіологічний стан людини за індивідуальними показниками співвідношень рівнів фізичного, емоційного та інтелектуального тону.

1.1. Метод дослідження психофізіологічного стану на основі кольорової преференції

В ході проходження тесту кольорової преференції, для визначення індивідуальних психофізіологічних особливостей, використовується комплект тестових карток еталонних кольорів, які представлені в двох варіантах: «БІОКОЛОП-I» і «БІОКОЛОП-II» (Рис 1.1, 1.2).

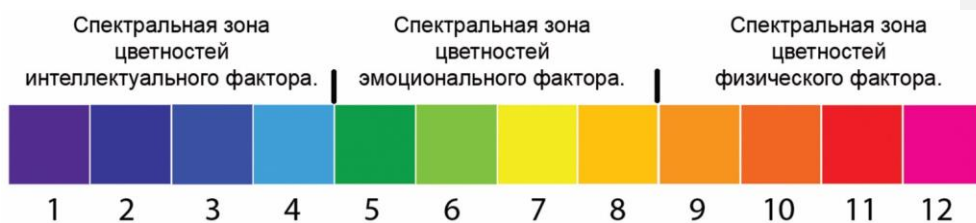


Рис 1.1 - Еталонні кольоровості тесту «БІОКОЛОП-I». Цифрами позначені номери карток

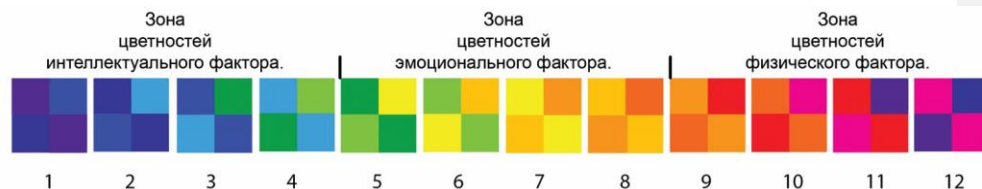


Рис 1.2 Еталонні кольоровості тесту «БІОКОЛОП-II». Цифрами позначені номери карток

Еталонні кольори тесту «БІОКОЛОП-I» та «БІОКОЛОП-II» розташовуються в спектральному ряду в порядку збільшення довжини електромагнітної хвилі.

Тильна сторона кольорових квадратів пронумерована відповідно до спектрального розташування кольорів за годинниковою стрілкою:

- фіолетова - 1, темно-синя - 2, синя - 3, блакитна - 4 – кольори інтелектуального фактору;
- зелена - 5, салатова - 6, яскраво-жовта - 7, тепло-жовта - 8 - кольори емоційного фактору;
- помаранчева - 9, оранжево-червона - 10, червона - 11, пурпурна - 12 - кольори фізичного фактору.

Використовувані в тесті кольори самооцінки фізичного, емоційного та інтелектуального тону стану людини, з точки зору психофізіології, відповідають певним суб'єктивним колірним асоціаціям [2].

У тестах «БІОКОЛОП-I» і «БІОКОЛОП-II» кольорові картки мають форму квадрата, оскільки людина сприймає квадрат в найбільш нейтральному поданні, в результаті чого мінімізуються асоціативні процеси і створюються умови для оцінювання колірної стимуляції в чистому вигляді, при цьому така форма рівномірно концентрує сприйняття характеристик кольору.

«БІОКОЛОП-II» являє собою 12 кольорових квадратних карток на яких кольори модифіковані шляхом парного діагонального ритмічного розташування кольорів на одній квадратній картці (згідно спектрального порядку). Таке розташування дозволяє підвищити рівень преференційного сприйняття, що відкриває можливості виявлення більш тонких особливостей індивідуальної преференції [1].

Для фіксування результату дослідження, проведеного за допомогою кольорових наборів «БІОКОЛОП-I» та «БІОКОЛОП-II», а також оцінка своїх суб'єктивних станів досліджуваного, використовуються спеціальні паперові бланки.

Стіл або поверхню, на якій будуть розкладатися картки при проведенні тесту, повинні бути нейтрального кольору. Приміщення для проведення тесту має бути добре освітлено, щоб легко розрізнялися всі 12 кольорів. Тест

проводиться індивідуально. На здійснення кольорово преференціального вибору досліджуваного надається не більше 30 секунд.

На антарктичній станції «Академік Вернадський» тест кольорової преференції являється частиною методики комплексного дослідження психофізіологічних особливостей людини, яка відбувається за алгоритмом (рис 1.3.)

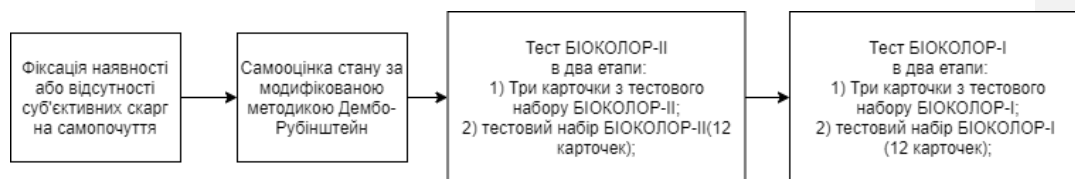


Рис 1.3 Порядок проведення комплексного дослідження психофізіологічних особливостей людини

Методика може застосовуватися і в скороченому варіанті і включати тільки один з двох тестових наборів «БІОКОЛОП». Проте застосування методики комплексно, з використанням двох тестових наборів «БІОКОЛОП-I» і «БІОКОЛОП-II», підвищує надійність отриманих результатів і необхідна у випадках наявності суб'єктивних скарг досліджуваних на самопочуття [1].

Група кольорів, обрана під час проходження тесту кольорової преференції першою, показує потребу в посиленні в першу чергу того фактору стану, до якого вона належить. Наступна за ними група кольорів показує, який з інших двох факторів може бути причиною ослаблення першого.

Якщо в процесі дослідження результату тестування було виявлено розбіжності в порядку розташування кольорових груп, під час проходження скороченого тесту (3 кольори) та повного (12 кольорів), це може бути ознакою зміни психофізіологічного стану. А розбіжності, виявлені при тестуванні двома

тестовими наборами, показують наявність дисбалансного стану досліджуваного, що вимагає більш пильної уваги.

За результатами дослідження надаються рекомендації щодо корекції психофізіологічного стану досліджуваних. Індивідуально і для групових тренінгів застосовуються кольороводинамічні картини-таблиці, які є складовою частиною методики і використовуються для релаксації, гармонізації та підвищення стійкості до екстремальних і стресових впливів, нормалізації проходження процесів адаптації в несприятливих умовах [3].

У процесі зорового відчуття і сприйняття кольорових картин-таблиць на тлі колірної адаптації проявляється ефект релаксації і виникають позитивні емоції, що важливо для корекції порушень психофізіологічного статусу людини. В основі біорегуляційної дії - активація механізмів відчуття і сприйняття закономірностей колірних контрастів [3]. Цей процес можна представити у вигляді такого собі, індивідуально обраного «візуального масажу» системи сенсорного сприйняття стимулів хвильового спектру в діапазоні видимого світла.

Результати дослідження а також дані зібрані під час експедиції - це тести кольорової преференції, суб'єктивні відчуття людини, можуть бути використані для прогнозування стану людини. Наразі такого прогнозування не виконувалось, проте такий підхід до дослідження дозволить виявляти ранні прояви погіршення психофізіологічного стану і прогнозувати погіршення самопочуття зимівників. Для прогнозування стану людини на основі цих даних доцільно використати методи машинного навчання.

1.2. Аналіз сучасних технологій обробки даних

Кажуть, що комп'ютерна програма вчиться з досвіду E по відношенню до якогось класу задач T та міри продуктивності P , якщо її продуктивність у задачах з T , вимірювана за допомогою P , покращується з досвідом E [4].

Машинне навчання - вузькоспеціалізована галузь знань, що входить до складу основних джерел технологій і методів, що застосовуються в областях великих даних і Інтернету речей, яка вивчає і розробляє алгоритми автоматизованого вилучення знань з сирого набору даних, навчання програмних систем на основі отриманих даних, генерації прогнозів або рекомендацій, розпізнавання образів і т.д.

Еволюціонувавши з досліджень розпізнавання образів та теорії обчислювального навчання в галузі штучного інтелекту, машинне навчання досліджує вивчення та побудову алгоритмів, які можуть навчатися й робити передбачення на основі даних [5].

Задачі машинного навчання в залежності від бажаного результату на виході системи поділяються на: задачі класифікації, задачі регресії, задачі кластеризації. За типом машинного навчання можна поділити на: навчання з учителем, навчання без вчителя, навчання з підкріпленням.

1.2.1. Задача класифікації

Задача класифікації – відноситься до розділу машинного навчання з учителем. Суть задачі зводиться до наступного. Існує масив об'єктів розділених за певною ознакою на класи. Задано кінцевий масив об'єктів для якого відомо, до якого класу відноситься об'єкт. Класова приналежність інших об'єктів невідома. Необхідно знайти алгоритм, здатний класифікувати випадковий об'єкт з початкового масиву.

Задача класифікації ділиться на бінарну класифікацію, коли об'єкт може належати до одного з двох класів, та багатокласової в якому об'єкт може належати до одного з множини класів, наприклад задача передбачення родовищ корисних копалин, по даним геологічної розвідки, виносить рішення про наявність або відсутність тієї чи іншої породи.

На рис 1.4 зображено типову бінарну класифікацію, в якому об'єкти позначені різними кольорами належать різним класам, і після роботи методу машинного навчання було знайдено алгоритм по якому в майбутньому будуть класифікуватись об'єкти.

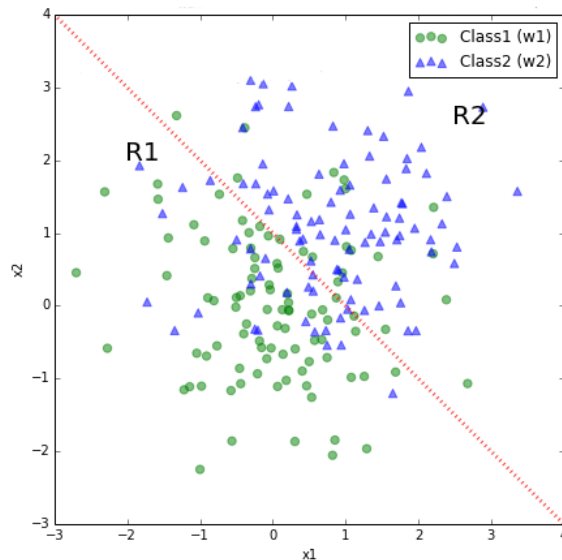


Рис 1.4 Приклад бінарної класифікації

Даний клас задач машинного навчання не зовсім підходить для прогнозування стану людини, оскільки вимагає щоб вихідними даними було чітко розділені класи, це можливо за умови що дані про суб'єктивні відчуття будуть зведені до одного параметру, в якому наприклад буде 3 суб'єктивні стани: поганий, нормальний, гарний.

1.2.2. Задача регресії

Задача регресії – також відноситься до машинного навчання з учителем. Суть задачі полягає в прогнозуванні вихідного значення Y , по відомому (вхідному) значенню X , тобто знаходження функції залежності Y від X ($f(X) = Y$). При цьому необхідно щоб існувала залежність між вхідними та вихідними даними.

Вихідним значення, на відміну від задачі класифікації буде являтися непереривною величиною, дійсне число – ціле або з плаваючою комою. Вхідне значення може бути одне, або декілька, в другому випадку йдеться про багатовимірну регресію.

На рис 1.5 зображено приклад роботи машинного навчання для вирішення задачі регресії, де точками позначено навчальні об'єкти, а лінія являється лінією регресії, яка показує залежність значення вихідного параметру від вхідного.

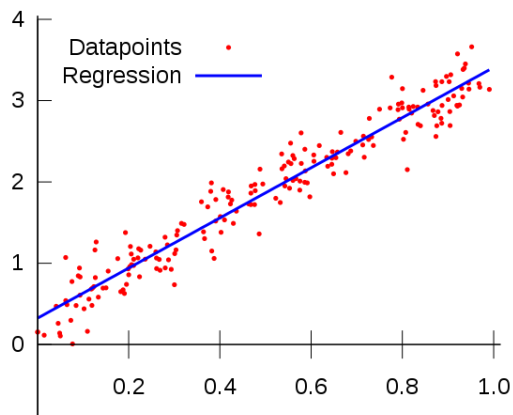


Рис 1.5 Приклад лінійної регресії

Підхід може бути використаний при умові що дані про суб'єктивний відчуття людини будуть зведені до одного параметру. При цьому даний підхід

дозволяє прогнозувати стан людини як неперервну величину, що дозволить отримувати більш точні результати прогнозування, ніж розбиття суб'єктивного стану на певні класи.

1.2.3. Задача кластеризації

Задача кластеризації – відноситься до машинного навчання без вчителя. Задача полягає в знаходженні деякого розбиття досліджуваної сукупності об'єктів на підмножини схожих між собою об'єктів. При цьому вихідним припущенням для виділення таких підмножин, що отримали спеціальну назву кластерів, служить лише неформальне припущення про те, що об'єкти, які відносяться до одного кластера, повинні мати більшу схожість між собою, чим з об'єктами з інших кластерів.

На рис 1.6 показано розподіл об'єктів на три групи (кластера) на основі залежностей між двома параметрами.

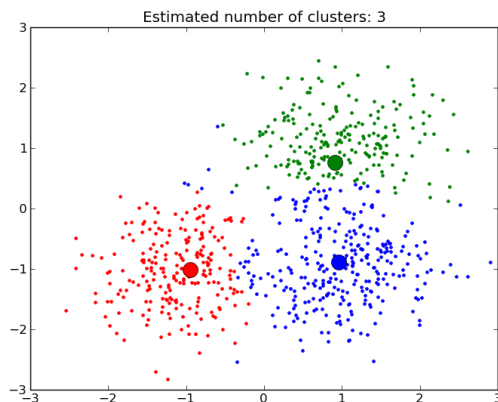


Рис 1.6 Приклад кластеризації з трьома кластерами

Відмінність класифікації та кластеризації полягає в тому, що при класифікації є набір визначених класів, навчання відбувається на наборі

прикладів. При кластеризації використовується алгоритм, який намагається згрупувати набір об'єктів і визначити, чи існує яка-небудь взаємозв'язок між об'єктами.

Розглянувши задачі, які може вирішувати машинне навчання, буде доцільним використати для розробки підходу машинне навчання з учителем, для вирішення задачі регресії, де в якості вхідних даних для тренування можна використати результати проходження тесту кольорової преференції, еталонне математично визначене значення результату проходження тесту кольорової преференції («БІОКОЛОР») та данні про суб'єктивній відчуття людини які були отримані від НАНЦ України. На основі даних про суб'єктивні відчуття людини необхідно вивести оцінку стану людини, дана величина є неперервною що і відповідає вихідним параметром задачі регресії, тому рішення задачі регресії буде спроектовано на прогнозування стану людини на основі проходження тесту кольорової преференції та «БІОКОЛОР».

Висновки:

1. Проведено огляд дослідження психофізіологічного стану людини на основі кольорової преференції.
2. Розглянуто задачі, які вирішує машинне навчання для обробки та аналізу даних із класифікацією за призначенням, типом вихідних даних, принципом роботи.
3. На основі проведеного огляду обрано задачу регресії для вирішення задачі прогнозування стану людини, з вхідними даними: тестами кольорової преференції, еталонного значення кольорової преференції, вихідні дані: оцінка стану людини, яка базується на суб'єктивних відчуттях. При такому підході можна дані про суб'єктивний відчуття людини звести до одного параметру який буде неперервною величиною.

Це дозволить більш точно прогнозувати стан людини, ніж розбиття суб'єктивного стану на певні класи.

РОЗДІЛ 2

ПІДХІД ДО АНАЛІЗУ ПСИХОФІЗІОЛОГІЧНОГО СТАНУ ЛЮДИНИ НА ОСНОВІ КОЛЬОРОВОЇ ПРЕФЕРЕНЦІЇ

Професійна діяльність в Антарктиці пов'язана з перебуванням людини в умовах ізоляції малого колективу, соціальної та сексуальної депривації, незвичайної активності регіональних природних факторів, що може негативно відбиватися на його психофізіологічному стані [1]. Відомо також, що зміна біологічного ритму нормального функціонування систем організму людини при його діяльності в Антарктиці відбивається не тільки на стані здоров'я, але і на його працездатності. При цьому в організмі виникають порушення на системному, органному і клітинному рівнях. Вони здатні модифікувати реакції функціональних систем на зовнішні і внутрішні подразнення.

На антарктичній станції фіксували велику кількість випадків порушень нормального режиму сну, симптомів головного болю, погіршення настрою і самопочуття. Перераховані ознаки спостерігалися майже у всіх членів експедиції (до 95%), в результаті чого зростав ризик порушення психоемоційної стабільності і взаємин в малому колективі.

Дослідження дозволить прогнозувати зміну суб'єктивного стану людини на основі проходження тесту кольорової преференції, що дає змогу завчасно реагувати на погіршення психофізіологічного стану людини.

2.1. Аналіз психофізіологічного стану людини на основі кольорової преференції.

Підхід до аналізу психофізіологічного стану людини на основі кольорової преференції дозволяє коректувати психофізіологічний стан людини шляхом підбору релаксуючих картинок, при цьому підбір картинок відбувається на

основі результату проходження тесту кольорової преференції, який відображає фізичну, емоційну та інтелектуальну складові психофізіологічного стану людини. Процес аналізу психофізіологічного стану людини на основі проходження тестів кольорової преференції складається з багатьох кроків:

- безпосереднє тестування, включає проходження тесту людиною, та запис даних;
- первинний аналіз результатів тестування (оцінка індивідуальної кольорової преференції шляхом нанесення на кругову діаграму (рис 2.1) [1];
- підбір релаксуючих картинок, які будуть коригувати психофізіологічний стан людини;
- аналіз отриманих даних вже безпосередньо після закінчення експедиції, знаходження закономірностей в даних;

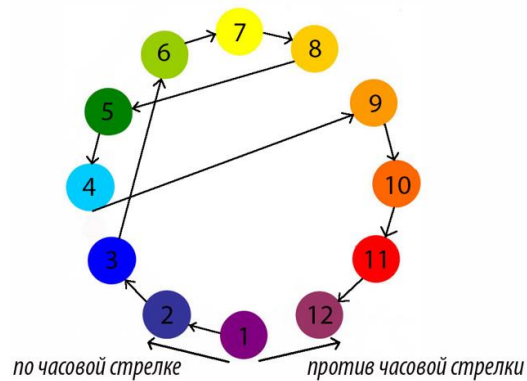


Рис 2.1 Приклад нанесення на кругову діаграму результату тесту
кольорової преференції

Даний підхід передбачає під час проходження тестування запис суб'єктивних відчуттів досліджуваного.

Добавлено примечание ((E51)): картинка не ваша, в тексте, где упоминается должна быть ссылка откуда взята, и так про все, что самостоятельно не писали/ не рисовали

Моніторинг результатів проходження тестів кольорової преференції зимівників можуть виявляти ранні прояви погіршення психофізіологічного стану і прогнозувати погіршення самопочуття зимівників. Здійснення цього моніторингу безпосередньо під час експедиції пов'язано з певними складнощами, оскільки потребує обробки великої кількості тестів кольорової преференції, що займає багато часу, тому для вирішення цієї проблеми необхідно розробити підхід до аналізу психофізіологічного стану людини на основі кольорової преференції, який здатен буде прогнозувати стан людини з метою раннього виявлення погіршення самопочуття і як наслідок погіршення психофізіологічного стану людини.

2.2. Еталонний результат кольорової преференції

На основі дати народження та дати розрахунку можливо побудувати графік, який надає математично розрахований результат нормального біологічного стану кольорової преференції (рис 2.2).

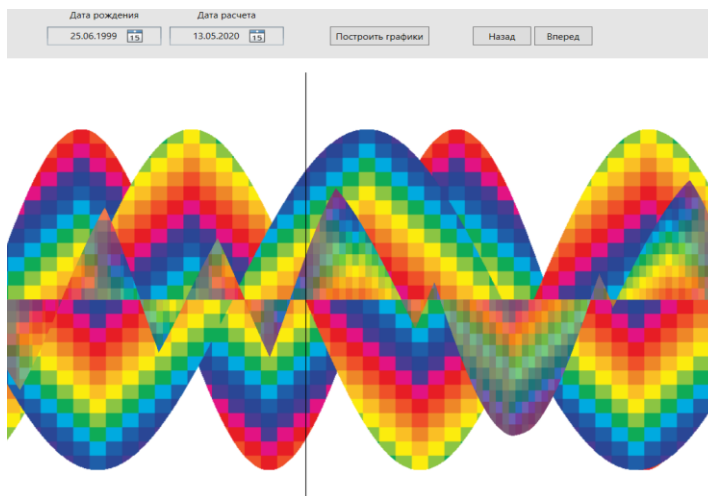


Рис 2.2 Приклад знаходження еталонного результату кольорової преференції.

2.3. Підхід до аналізу психофізіологічного стану людини на основі кольорової преференції та суб'єктивного відчуття людини

Необхідно спрогнозувати стан людини, подавши на вхід послідовність тесту преференцій та еталонне значення кольорової преференції та отримати на виході прогнозовану оцінку стану людини, яка базується на суб'єктивних відчуттях. Також необхідно підібрати найкращий метод машинного навчання, який буде мати найменшу вірогідність помилки та не буде потребувати масивних математичних розрахунків.

Дано:

- результати тестів «БЮКОЛОП–І» та «БЮКОЛОП–ІІ»;
- еталонне значення кольорової преференції;
- оцінка суб'єктивного стану зимівника;

Знайти:

- прогнозований стан зимівника;

НАНЦ України надав набір даних з пройденими тестам кольорової преференції, а також данні про суб'єктивного відчуття зимівників на основі яких можливо вирішити поставлену задачу.

На рис 2.3 зображено складові частини підходу для реалізації запропонованого підходу до аналізу психофізіологічного стану людини на основі кольорової преференції та суб'єктивних відчуттів людини.



Рис 2.3 Складові частини запропонованого підходу

На першому етапі необхідно розрахувати еталонний результат кольорової преференції. Розрахунок відбувається на основі дати народження людини, яка проходила тест кольорової преференції, та дати проведення тесту. В результаті отримаємо одну з чотирьох: жовту, синю, червону або змішану результуючу кольорову преференцію.

На наступному етапі для оцінки стану людини було надано данні про суб'єктивні відчуття зимівників які включали в себе такі пункти: головний біль, поганий сон, поганий настрій, загальна кваліість, поганий апетит, також були наданні данні про звернення до лікаря та тривалість лікування захворювань в період перебування на арктичній станції. Оцінка стану людини розглядалась як сума всіх факторів з певним коефіцієнтом, для негативних суб'єктивних відчуттів коефіцієнт = 1 , для звернень до лікаря = 10. Дана оцінка може приймати значення $[0 , \infty)$, 0 – стан без будь яких скарг на самопочуття(відсутність всіх негативних суб'єктивних відчуттів), чим більше

значення, тим гірший стан людини. Оцінка стану людини використовується як вихідне значення y .

Після цього необхідно виконати фільтрацію даних, яка включає в себе: видалення даних в яких відсутній якийсь з параметрів, перевірку значень тестів кольорової преференції – повинно бути цілими числами від 1 до 12, еталонний результат кольорової преференції - може приймати значення цілого числа від 1 до 4, також відбувається нормалізація оцінки стану людини.

Для прогнозування стану людини використовується метод машинного навчання, де на основі вхідних значень x – це послідовності карток «БІОКОЛОР» (I та II), та еталонні значення кольорової преференції

$$\{x_1, \dots, x_n\} \subset X,$$

де n – кількість тестів преференцій в наборі тестових даних;

відбувається прогнозування вихідного значення y - оцінка стану людини:

$$y_i = y(x_i), \quad y > 0;$$

Для того щоб оцінка стану людини була якомога краща, необхідно провести дослідження ефективності методів машинного навчання, які вирішують задачу регресії для виявлення кращих методів для прогнозування стану людини шляхом порівняння результату їх роботи. Після чого кращий з методів машинного навчання буде застосований для прогнозування стану людини.

Висновки:

1. У розділі описано проблеми дослідження психофізіологічного стану людини на основі кольорових преференцій, які пов'язані із ручним методом визначення кольорової преференції, еталонне значення результату кольорової преференції.

2. Поставлено завдання розробки підходу до аналізу психофізіологічного стану людини на основі тесту кольорової преференції та суб'єктивних відчуттів, який здатен давати оцінку стану людини.

3. Запропоновано підхід до аналізу психофізіологічного стану людини на основі тесту кольорової преференції та суб'єктивних відчуттів, який здатен давати оцінку стану людини.

РОЗДІЛ 3

ДОСЛІДЖЕННЯ ЕФЕКТИВНОСТІ МЕТОДІВ МАШИННОГО НАВЧАННЯ ДЛЯ ПРОГНОЗУВАННЯ СТАНУ ЛЮДИНИ.

У попередньому розділі було визначено що для прогнозування стану людини буде доречно використати машинне навчання для вирішення задачі прогнозування стану людини. Також необхідно провести аналіз та дослідження методів машинного навчання для вирішення задачі регресії для виявлення кращих методів для прогнозування стану людини, шляхом порівняння коефіцієнтів детермінації, середньоквадратичної помилка, середньої абсолютної помилки, а також враховувати швидкодію методів, їх переваги та недоліки.

3.1. Лінійна регресія

Для оцінки стану людини на основі тесту кольорової преференції та еталонного значення кольорової преференції використаємо модель множинної лінійної регресії (MLR) (рис 3.1). Перевагою лінійної регресія є швидкість і простота отримання моделі, при цьому дійсна модель є прозорою і зрозумілою для аналітика. За отриманими коефіцієнтами регресії можна давати оцінку наскільки той чи інший фактор впливає на результат, зробити на цій основі додаткові корисні висновки. Велика кількість реальних процесів можна з достатньою точністю описати лінійними моделями. Однак головним і суттєвим недоліком являється неможливість застосування даної регресії коли існує складна залежність між вхідними та вихідними значеннями, або в випадку не лінійності цієї залежності.

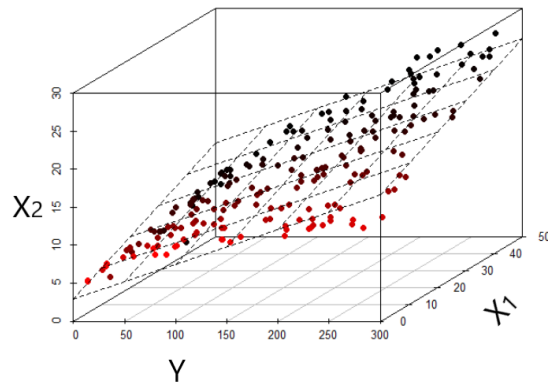


Рис 3.1 Приклад регресивної моделі для множинної лінійної регресії з двома вхідними величинами X_1 , X_2

Рівняння, що описує зв'язок прогнозованого значення від вхідних величин, відоме як регресійна модель. Множинна лінійна регресія відрізняється від простої лінійної регресії тим що регресивною моделлю являється гіперплощина, а не проста лінія регресії.

Рівняння регресії (формула 3.1):

$$\hat{y}(w, x) = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p$$

Формула 3.1 – Рівняння регресії

\hat{y} – прогнозоване значення;

β_1, \dots, β_p - коефіцієнти функції лінійної регресії;

β_0 - незалежний член;

x_1, \dots, x_p - еталонне значення кольорової преференції, та дані тесту кольорової преференції;

В множинній лінійній регресії можливий прояв мультиколінеарності. Наявність високої кореляції між вхідними змінними моделі множинної лінійної регресії, що може викликати нестійкість роботи моделі, коли коефіцієнти регресії сильно змінюються навіть при незначній зміні вихідних даних.

Зростання дисперсії оцінки коефіцієнтів регресії веде до того, що навіть схожі дані, що подаються на вхід моделі, можуть привести до різної оцінки і навіть суперечливих результатів. Для боротьби з мультиколінеарністю використовуються різні методи, такі як виняток сильно корельованих змінних, комбінування змінних, метод головних компонент і ін.

Для вирішення задачі прогнозування стану людини було використано лінійну регресію в якій коефіцієнти моделі множинної лінійної регресії, знаходять за допомогою методу найменших квадратів. Метод найменших квадратів дозволяє знайти такі значення коефіцієнтів, що сума квадратів відхилень буде мінімальною. Розглянемо систему рівнянь з n лінійних рівнянь, з невідомими коефіцієнтами, $\beta_1, \beta_2, \dots, \beta_p$, при $n > p$ (формула 3.2) :

$$\sum_{j=1}^p X_{ij} \beta_j = y_i, \quad (i = 1, 2, \dots, n)$$

Формула 3.2 – Система рівнянь для знаходження невідомих коефіцієнтів рівняння регресії

В матричному вигляді буде: $X\beta = y$, де

$$X = \begin{bmatrix} X_{11} & X_{12} & \cdots & X_{1p} \\ X_{21} & X_{22} & \cdots & X_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ X_{n1} & X_{n2} & \cdots & X_{np} \end{bmatrix}, \quad \beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{bmatrix}, \quad y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}.$$

Для знаходження коефіцієнтів використається формула 3.3

$$\hat{\beta} = (X^T X)^{-1} X^T y$$

Формула 3.3 – Знаходження коефіцієнтів множинної лінійної регресії в матричному виді

Після проведення навчання моделі для прогнозування оцінки стану людини було отримано наступні коефіцієнти:

$$\hat{\beta} = \begin{bmatrix} -0.37761408 \\ -0.78184267 \\ -0.11692489 \\ -0.46884104 \\ -0.1643032 \\ -0.4176863 \\ -0.07954877 \\ -0.4750741 \\ -0.28023735 \\ -0.25892246 \\ -0.49450128 \\ -0.25013087 \\ -0.37961796 \end{bmatrix}$$

Незалежний член: $\beta_0 = 30.14845861$

Для оцінки ефективності методів навчання будемо використовувати наступні метрики:

- коефіцієнт детермінації: 0.224 ;
- середньоквадратична помилка: 4.497 ;
- середня абсолютна помилка: 1.656;

3.2. Регресія опорних векторів

Метод опорних векторів (SVM) - це універсальна модель машинного навчання, здатна виконувати лінійну або нелінійну класифікацію, регресію, виявлення викидів. Вона є однією з найпопулярніших моделей в машинному навчанні.

Алгоритм SVM досить універсальний: він підтримує не тільки лінійну і нелінійну класифікацію, але також лінійну і нелінійну регресію. Прийом при якому метод опорних векторів використовують для вирішення задачі регресії полягає в інвертуванні мети: замість спроби пристосуватися до найширшої з можливих смуг між двома класами, одночасно обмежуючи порушення зазору, регресія SVM намагається вмістити якомога більше зразків на смузі поряд з обмеженням порушень зазору (тобто зразків поза смуги), даний метод і називається регресією опорних векторів.

Ширина смуги зазору змінюється параметром C . На рис 3.2 показані дві моделі регресії SVM, навчені на випадкових даних, одна з широким зазором ($C = 0,01$) і одна з вузьким зазором ($C=100$). Якщо додавання додаткових навчальних зразків всередину зазору не впливає на прогнози моделі, можна зробити висновок що модель нечутлива до C .

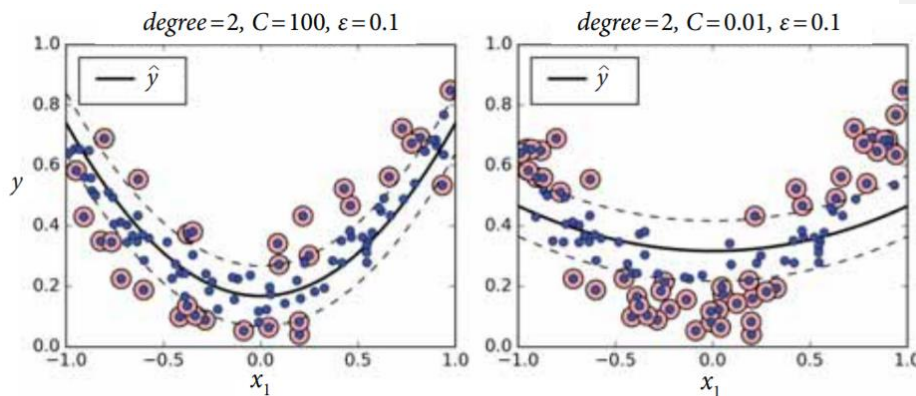


Рис 3.2 Приклад регресії опорних векторів з поліноміальним ядром з різними значеннями параметру C .

При наявності нелінійного зв'язку між вхідними та вихідними величинами якість лінійних класифікаторів, часто може виявитися незадовільним. Для врахування нелінійності зазвичай розширюють простір змінних, включаючи

різні функціональні перетворення вихідних предикатів (поліноми, експоненти і ін.). SVM - можна розглядати як нелінійний узагальнення лінійного класифікатору, основане на розширенні розмірності вхідного простору предикатів за допомогою спеціальних ядерних функцій. Це дозволяє будувати моделі з використанням поверхонь різної форми.

Ядром є будь-яка симетрична, позитивно напіввизначена матриця K , яка складена з скалярних добутку пар векторів x_i і x_j : $K(x_i, x_j) = (\varphi(x_i), \varphi(x_j))$, яка характеризує міру їх близькості. φ - довільна перетворююча функція, що формує ядро. В якості таких функцій для прогнозування оцінки стану людини на основі тесту кольорової преференції було розглянуто та використано наступні ядерні функції:

- лінійну: $\langle x, x' \rangle$
- поліномну: $(\gamma \langle x, x' \rangle + r)^d$, де d - параметра degree.
- радіально базисну функцію (RBF): $\exp(-\gamma \|x - x'\|^2)$, де γ визначено параметром gamma, має бути більше 0.

Для того щоб модель мала меншу похибку необхідно попередньо підібрати значення ряду параметрів. Параметр C , загальний для всіх ядер SVM, при вирішенні задачі класифікації являє собою допустимий штраф за порушення кордону, для задачі регресії - штраф за вихід за межі кордонів.

Значення параметрів для ядер буде впливати на якість оцінки суб'єктивного відчуття, так було розглянуто та визначено найкращі параметри для даного завдання (рис 3.3, рис 3.4, рис 3.5, рис 3.6).

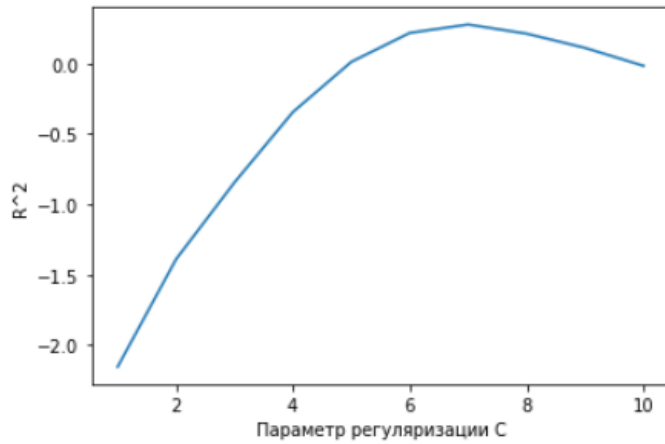


Рис 3.3 Залежність між параметром C та коефіцієнтом детермінації моделі навчений методом опорних векторів з використанням поліномної функції в якості ядра.

d (degree) - степінь функції ядра полінома

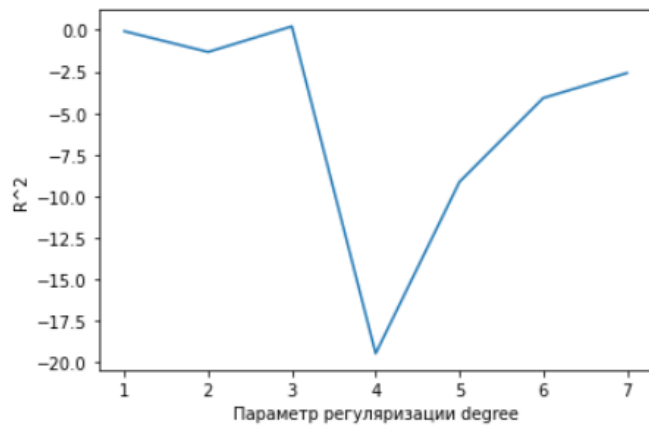


Рис 3.4 Залежність між параметром d та коефіцієнтом детермінації моделі навчений методом опорних векторів з використанням поліномної функції в якості ядра.

Для моделі з поліномного ядра $\gamma\langle x, x' \rangle + r)^d$ найкраща дисперсія буде при $C = 7$ та $d = 3$, подальше збільшення d являється не актуальним оскільки потребує дуже великих розрахунків, що займає багато часу, при цьому надає гірші результати.

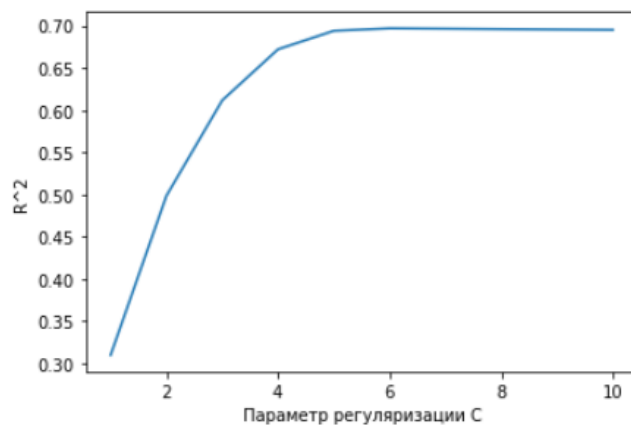


Рис 3.5 Залежність між параметром C та коефіцієнтом детермінації моделі навчений методом опорних векторів з використанням радіальної базисної функції в якості ядра.

γ (gamma)- параметр радіальної функції - визначає, який вплив надає окремий навчальний приклад. Чим більше γ , тим більше інші приклади повинні бути порушені.

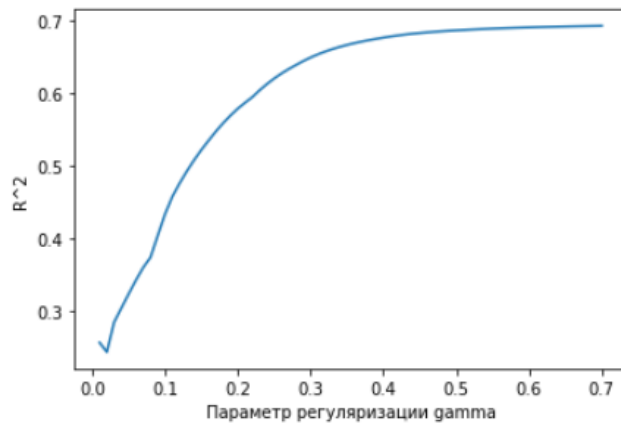


Рис 3.6 Залежність між параметром γ та коефіцієнтом детермінації моделі навчений методом опорних векторів з використанням радіальної базисної функції в якості ядра.

Для радіальної базисної функції $\exp(-\gamma \|x - x'\|^2)$: найкраща дисперсія буде при $C = 9$, та $\gamma = 0,6$, подальше збільшення параметрів не призводить до покращення моделі.

Метод опорних векторів ефективний при великій кількості вхідних значень, в тих випадках, коли кількість вимірювань перевищує кількість зразків. Великою перевагою даного метода є універсальність: різні функції ядра можуть бути вказані для вирішальної функції. Надаються загальні ядра, але також можливе створення власних функцій які будуть використовуватись для ядра.

До недоліків даного метода можна віднести:

- необхідність підбору параметрів для покращення роботи моделі;
- швидкодія методу, оскільки деякі з ядер можуть мати великі розрахунки на які потрібно багато часу;

Для оцінки стану людини було використано метод опорних векторів з поліномною та радіально базисною функцією ядра. При цьому для оцінки ефективності роботи методу для різних ядер було підбрано параметри які дають найкраще значення коефіцієнту детермінації.

Для методу з радіально базисною функцією ядра:

- коефіцієнт детермінації: 0.644;
- середньоквадратична помилка: 3.282;
- середня абсолютна помилка: 1.365;

Для методу з поліномною функцією ядра:

- коефіцієнт детермінації: 0.279;
- середньоквадратична помилка: 6.047;
- середня абсолютна помилка: 1.075;

3.3. Метод k-найближчих сусідів

Моделі навчені методом k-найближчих сусідів (KNN) прості в реалізації і добре справляються з нелінійностями. Регресія на основі методу k-найближчих сусідів може використовуватися в тих випадках, коли дані на виході моделі є безперервними, а не дискретними змінними. А вихідне значення, обчислюється на основі середнього значення міток її найближчих сусідів.

Різні атрибути можуть мати різний діапазон представлених значень у вибірці (наприклад атрибут А представлений в діапазоні від 0,1 до 0,5, а атрибут Б представлений в діапазоні від 1000 до 5000), то значення дистанції можуть сильно залежати від атрибутів з великими діапазонами. Тому дані зазвичай підлягають нормалізації. У зв'язку з цим необхідно перед початком навчання моделі виконати нормалізацію даних.

Прогнозувати на основі методу k-найближчих сусідів повільніше порівняно з іншими методами, оскільки для прогнозування результату

необхідно буде виконати пошук по всіх точках в своєму навчальному наборі, щоб знайти найближчі. Тому для великих наборів даних KNN буде повільним методом в порівнянні з іншими регресіями, для адаптації яких може знадобитися більше часу, але потім роблять свої прогнози за допомогою простих обчислень.

Ще одна проблема з моделлю KNN полягає в тому, що вона погано інтерпретується. Лінійна регресія матиме чітко інтерпретовані коефіцієнти, які самі можуть дати певне уявлення про вплив даної ознаки на вихідне значення. Однак знаходження які саме функції мають найбільший ефект, не має сенсу для моделі KNN. Частково через це моделі KNN також не можуть бути використані для вибору ознак, як може бути лінійна регресія з додатковим терміном функції вартості або спосіб, яким дерево рішень неявно вибирає один найбільш цінний параметр.

Для використання методу k-найближчих сусідів, необхідно визначитися з певною кількістю сусідів(K), метрикою відстані, та вагою найближчих сусідів.

Збільшення кількості сусідів матиме тенденцію покращувати результат роботи моделі до певного значення. Не існує єдиного значення k, яке буде працювати для кожного набору даних. Для моделей регресії, особливо якщо є тільки два класи, зазвичай вибирається непарне число для k. Це зроблено з цією метою, щоб алгоритм ніколи не стикається з випадком коли при виборі чотирьох найближчих точок, дві з них знаходяться в одному класі, а дві - в іншому.

Для визначення оцінки суб'єктивного стану було розглянуто моделі з параметром K від 1 до 15, в результаті отримано що найкращий коефіцієнт детермінації при K=3 (рис 3.7).

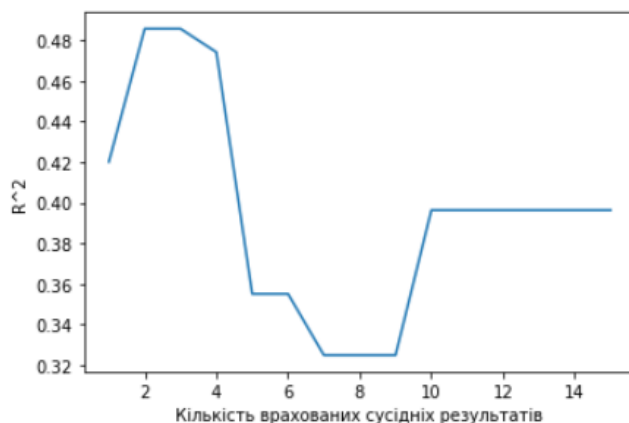


Рис 3.7 Залежність між кількістю сусідів(k) та коефіцієнтом детермінації моделі.

Способи вимірювання того, наскільки «близькі» дві точки одна від одної, і відмінності між методами пошуку відстані між сусідами можуть мати вагомий вплив на якість моделі. Найчастіше використовується евклідова відстань. Інший спосіб вимірювання - це так звана «відстань Манхеттена», при якій вимірюється відстань, взята в кожному кардинальному напрямку, а не по діагоналі (як ніби ви йшли від одного перехрестя вулиць в Манхеттені до іншого і повинні були слідувати за сіткою вулиць, а не бути можливість вибрати найкоротший маршрут «по прямій лінії») [7]. У більш загальному сенсі, це насправді обидві форми так званого «відстані Маньківського» (формула 3.4):

$$d(X, y) = \left(\sum_{i=1}^n |x_i - y_i|^p \right)^{1/p}$$

Формула 3.4 - «відстань Маньківського»

Коли p встановлено в 1, ця формула збігається з манхеттенською відстанню, а коли встановлено в два - евклідовою відстанню.

Один із способів вирішити проблему можливого «зв'язування», коли наші прогнози регресії погіршилися по відношенню до кордонів набору даних, полягає у використанні зваженого способу (рис 3.8).

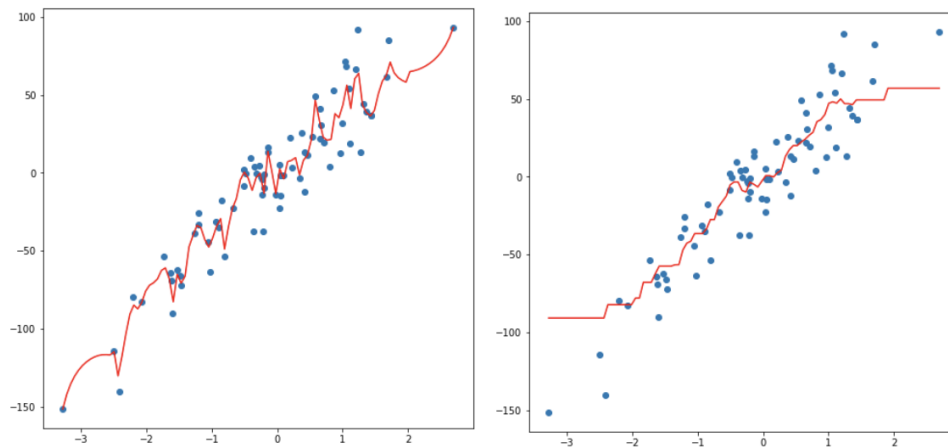


Рис 3.8 Вплив на функцію регресії зваженого (зліва) та не зваженого (справа) способів.

При такому способі, найближчі точки будуть враховуватися з більшим коефіцієнтом, ніж точки далі. Алгоритм буде як і раніше дивитися на всіх k найближчих сусідів, але ближчі сусіди матимуть більший вплив, ніж ті, які знаходяться далі. Для K найближчих сусідів буде визначатись оцінки близькості:

$$Q_j = \sum_{i=1}^n \frac{1}{d(x, a_i)^2}$$

Де $d(x, a_i)^2$ відстань від нового значення x до об'єкта a .

В таблиці 3.1 наведені різні варіанти параметрів для навчання моделі методом к-найближчих сусідів.

Таблиця 3.1

Вплив використання різних способів вимірювання відстані між сусідами та зваженого/незваженого способів на модель для прогнозування стану людини

	Евклідова відстань/без зважування	Евклідова відстань/з зважуванням	відстань Манхеттена /без зважування	відстань Манхеттена /з зважуванням
Коефіцієнт детермінації	0.016	0.420	0.006	0.420
Середньоквадратична помилка	13.455	5.091	13.227	5.091
Середня абсолютна помилка	2.273	1.000	2.227	1.000

Як видно з таблиці 3.1 метрика відстані не значно впливає на результати роботи моделі, при врахуванні ваги сусідів значно покращується робота моделі. Тому доцільно буде використовувати модель з врахуванням ваги сусідів, «відстанню Манхеттен», оскільки вона швидше розраховується, з кількістю сусідів для врахування $K=3$.

3.4. Регресійне дерево

Дерева рішень здійснюють розбиття простору об'єктів відповідно до деяких наборів правил розбиття. Ці правила є логічними твердженнями щодо тієї чи іншої змінної і можуть бути істинними або помилковими. Ключовими тут є три обставини:

- правила дозволяють реалізувати послідовну дихотомічну сегментацію даних;

- два об'єкти вважаються схожими, якщо вони виявляються в одному і тому ж сегменті розбиття;
- на кожному кроці розбиття збільшується кількість інформації щодо досліджуваної змінної.

Щоб досягти хороших результатів при вирішенні задачі регресії, необхідно щоб дерево рішень мало значно більші розміри ніж при вирішенні задачі класифікації.

Дерева класифікації і регресії є одним з найбільш популярних методів вирішення багатьох практичних завдань, що обумовлено наступними причинами:

- дерева рішень дозволяють отримувати легко інтерпретуємі моделі, що представляють собою набір правил виду "якщо ..., то ...". Інтерпретація полегшується, в тому числі, за рахунок можливості подати ці правила у вигляді деревовидної структури;
- в силу своєї роботи дерева рішень дозволяють працювати зі змінними будь-якого типу, без необхідності будь-якої попередньої підготовки цих змінних, для введення в модель (наприклад, логарифмування, перетворення категоріальних змінних в індикаторні, і т.п.);
- дерева рішень, по суті, автоматично виконують відбір інформативних аргументів і враховують можливі взаємодії між ними. Це, зокрема, робить дерева рішень корисним інструментом розвідувального аналізу даних;
- дерева рішень однаково добре застосовні як до кількісних, так і до якісних залежних змінних;

Недоліками даних методів є нестабільність і невисока точність прогнозів, що, не завжди підтверджується. За своєю суттю, дерева використовують "наївний підхід" в тому сенсі, що вони виходять із припущення про взаємну

незалежності ознак [7]. Тому моделі регресійних дерев статистично найбільш працездатні, коли комплекс аналізованих змінних є не надто мультиколінійності або є регулярна внутрішня множинна альтернатива у вихідній комбінації ознак.

Оскільки тест кольорової преференції являє собою певну послідовність з 12 карток, а також знаючи що еталонне значення результату кольорової преференції може мати всього 4 стани, можна припустити що для цих даних необхідна глибина дерева приблизно 135.

Під час підбору найкращої максимальної глибини дерева (рис 2.10, 2.11), було виявлено що на даному наборі даних метод працює дуже не стабільного, і тому використання його для прогнозування стану людини є недоцільним.

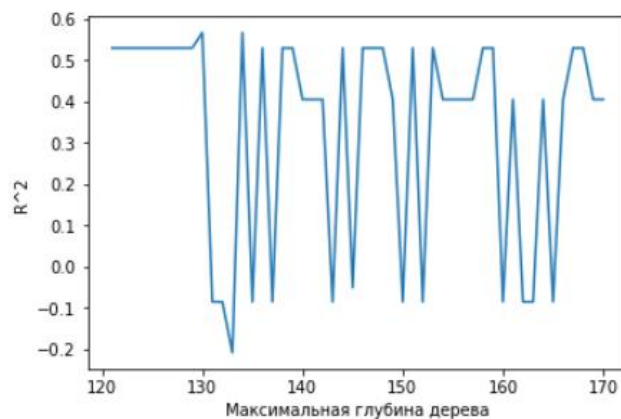


Рис 3.9 Залежність між максимальною глибиною дерева та коефіцієнтом детермінації моделі.

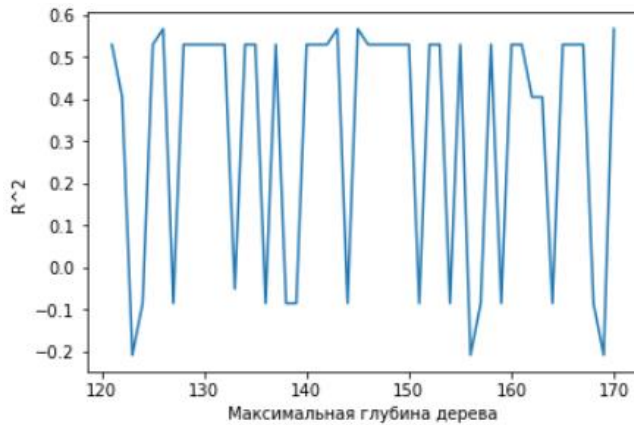


Рис 3.10 Залежність між максимальною глибиною дерева та коефіцієнтом детермінації моделі.

Для вирішення проблеми нестабільності методу регресійного дерева, використовують ансамбль випадкових дерев. В даному випадку для вирішення нестабільності використовується метод екстра випадкових дерев. У звичайних випадкових деревах випадковість випробовується на один шаг вперед в способі розрахунку розбиття. Як і в випадкових лісах, використовується випадкова підмножина об'єктів-кандидатів, але замість пошуку найбільш дискримінаційних порогів, пороги мають випадковий режим для кожного об'єкта-кандидата, і найкращий з цих випадково генерованих порогів вибирається в якості правила розділення. Це зазвичай дозволяє зменшити дисперсію за рахунок збільшення зміщення.

Для методу екстра випадкових дерев було проведено 20 тестів (рис 2.12) навчання моделі, при цьому дисперсія знаходилась в межах від 0,53 - 0,59, з чого можна вважати, модель є стабільною.

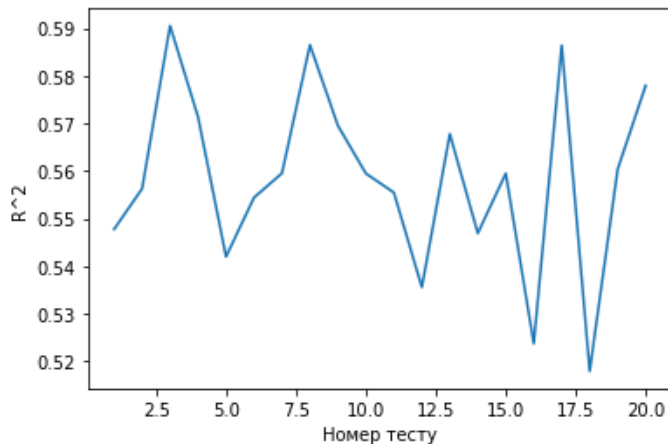


Рис 3.11 Перевірка стабільності моделі навченій методом екстра випадкових дерев.

Проблема з кількома виходами - це контрольоване завдання навчання, з декількома виходами для прогнозування, тобто, коли Y - це n -вимірний масив.

Коли немає кореляції між вихідними даними, дуже простий спосіб вирішити цю проблему - побудувати n незалежних моделей, тобто одну для кожного вихідного сигналу, а потім використовувати ці моделі для незалежного прогнозування кожного з n вихідних даних. Однак, оскільки цілком імовірно, що вихідні значення, що відносяться до одного і того ж вхідного набору даних, самі корелюють, і тому найкращим способом буде створення єдиної моделі, здатної прогнозувати одночасно всі n вихідних даних. Це вимагає меншого часу навчання, оскільки будується тільки одна модель. Точність узагальнення отриманої оцінки часто може бути збільшена.

Дерев рішень, можуть бути легко використані для вирішення проблем з декількома виходами. Це вимагає наступних змін:

- замість однієї вихідної величини необхідно використати n - величин;

- використання критерію поділу, які обчислюють середнє зменшення за всіма n вихідних даних.

Дану перевагу можна застосувати для того щоб не просто прогнозувати оцінку стану людини, а прогнозувати які саме негативні ефекти будуть проявлятися в людини. Вихідними параметрами будуть виступати негативні суб'єктивні відчуття які були надані НАНЦ України.

В результаті для перевірки коректності навчання моделі, та для уникнення нестабільності моделі було проведено 20 тестів (рис 2.13) навчання моделі, при цьому дисперсія знаходилась в межах від 0,48 - 0,57, що можна вважати стабільною моделлю.

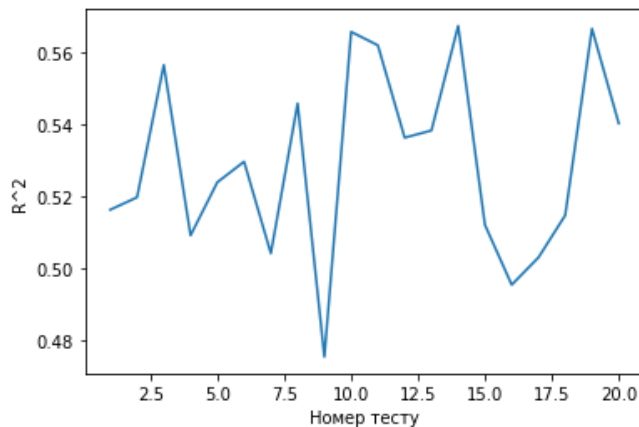


Рис 3.12 Перевірка стабільності моделі з n - виходами.

В результаті виявлено що помилка прогнозування буде більша при прогнозуванні конкретних негативних факторів ніж оцінки стану людини при навчанні моделі на основі ансамблю дерева рішень.

Під час застосування методу регресійного дерева для навчання моделі виявлено, що модель є нестабільною, для уникнення цього був застосований

один з методів ансамблю дерев. Модель навчена методом екстра випадкових дерев є стабільною та має наступні метрики:

- коефіцієнт детермінації: 0.644;
- середньоквадратична помилка: 3.282;
- середня абсолютна помилка: 1.365;

3.3. Порівняння ефективності методів машинного навчання для прогнозування стану людини.

В результаті отриманих метрик для оцінки ефективності методів машинного навчання можна підібрати найкращий метод для прогнозування оцінки стану людини. Для наглядного порівняння методів представимо параметри оцінки різних моделей в вигляді гістограм(рис 3.2, рис 3.3, рис 3.4).

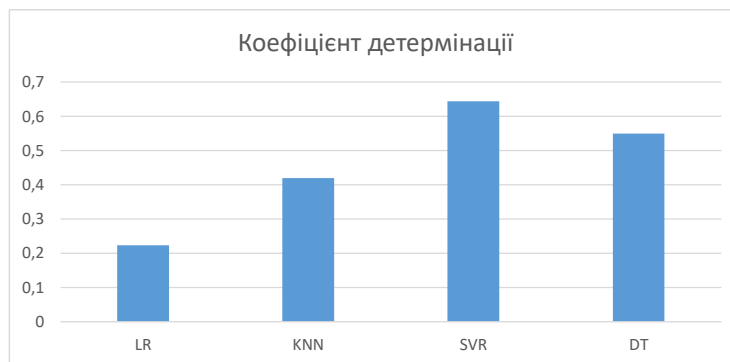


Рис 3.13 Коефіцієнти детермінації для моделей на основі різних методів машинного навчання.

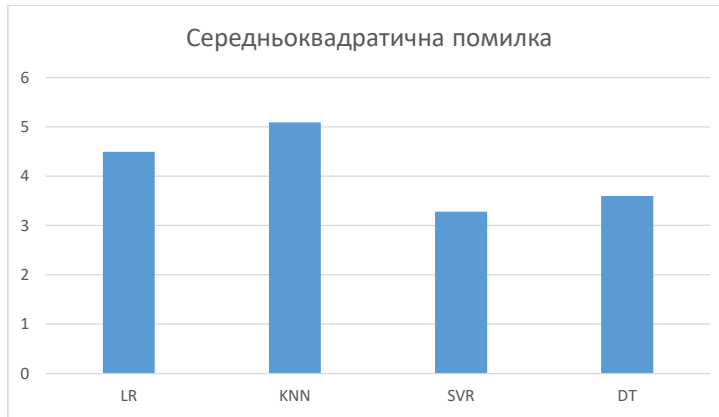


Рис 3.14 Середньоквадратична помилка для моделей на основі різних методів машинного навчання.

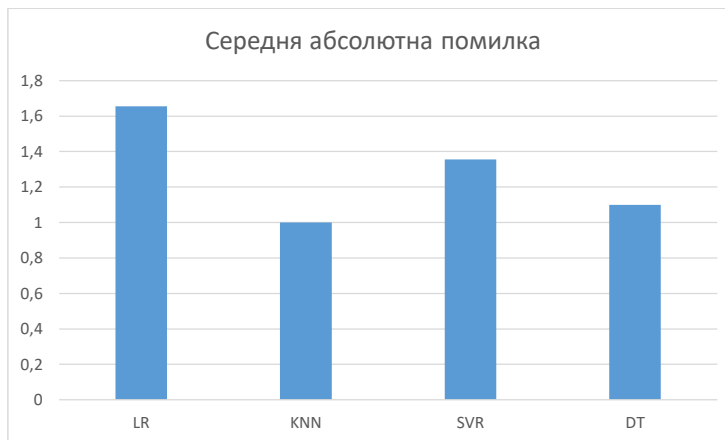


Рис 3.15 Середня абсолютна помилка для моделей на основі різних методів машинного навчання.

Як видно з гістограм найкращим методом для навчання моделі виявився метод опорних векторів оскільки має найкращі коефіцієнт детермінації та середньоквадратичну помилку. В якості ядра для даного методу

використовувалась радіально базисна функція виду $\exp(-\gamma \|x - x'\|^2)$ з $\gamma = 0,6$.

Недоліком даного методу являється те що потрібно підбирати параметри для моделі, та проводити нормалізацію даних які подаються на вхід моделі.

В модель на основі методу екстра випадкових дерев відсутні ці негативні фактори, однак і параметри оцінки моделі нижчі. Моделі на основі лінійної регресії та методу к-найближчих сусідів виявились не здатними ефективно прогнозувати оцінку стану людини.

Тому для програмного забезпечення, яке реалізує підхід до аналізу психофізіологічного стану людини на основі кольорової преференції та суб'єктивних відчуттів буде використаний метод опорних векторів з радіально базисною функцією виду $\exp(-\gamma \|x - x'\|^2)$ з $\gamma = 0,6$.

Висновки:

1. Виявлено, що методи лінійної регресії та к-найближчих сусідів не здатні ефективно прогнозувати стан людини на основі тестів кольорової преференції та еталонного значення кольорової преференції.
2. Модель навчена на основі методу регресійного дерева виявились нестабільною, тому для вирішення поставленого завдання був використаний один з методів ансамблю дерев – екстра випадкових дерев.
3. Найкращим методом для прогнозування стану людини є метод опорних векторів, оскільки має найкращі коефіцієнт детермінації та середньоквадратичну помилку, тому даний метод і буде використовуватися для реалізації програмного забезпечення.

РОЗДІЛ 4

РЕАЛІЗАЦІЯ ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ ЗАПРОПОНОВАНОГО ПІДХОДУ ДО АНАЛІЗУ ПСИХОФІЗІОЛОГІЧНОГО СТАНУ ЛЮДИНИ НА ОСНОВІ КОЛЬОРОВОЇ ПРЕФЕРЕНЦІЇ ТА СУБ'ЄКТИВНИХ ВІДЧУТІВ.

У розділі розглядається опис програмної реалізації підходу до аналізу психофізіологічного стану людини. Результат застосування підходу на реальних даних отриманих від НАНЦ України. Вхідні та вихідні параметри для програмної реалізації підходу

4.1. Постанова завдання та вимоги до ПЗ

Розроблене ПЗ повинне:

- надавати можливість прогнозувати оцінку стану зимівника на основі пройденого тесту кольорової преференції та еталонного результату кольорової преференції;

Створене програмне забезпечення буде являтися окремим модулем, що дозволить використовувати його в інших програмах, в перспективі здійснити інтеграцію в ПЗ НАНЦ України. Також такий підхід дозволить вносити зміни, покращувати та виправляти помилки ПЗ без втручання в ПЗ в якому був використаний даний модуль.

4.2. Опис розробленого ПЗ

На рис 4.1 зображено послідовність дій розробленого підходу до аналізу психофізіологічного стану людини. Функція `GetDataFrame` повертає дані про проходження тестів преференції, оцінки суб'єктивного стану зимівників, еталонні результати кольорової преференції в форматі `DataFrame` з файлу по

вказаній назві, допустимі формати файлу csv, xlsx, xls. Для правильної роботи необхідно, щоб дані, які знаходяться в файлі, були розміщені згідно певного шаблону.

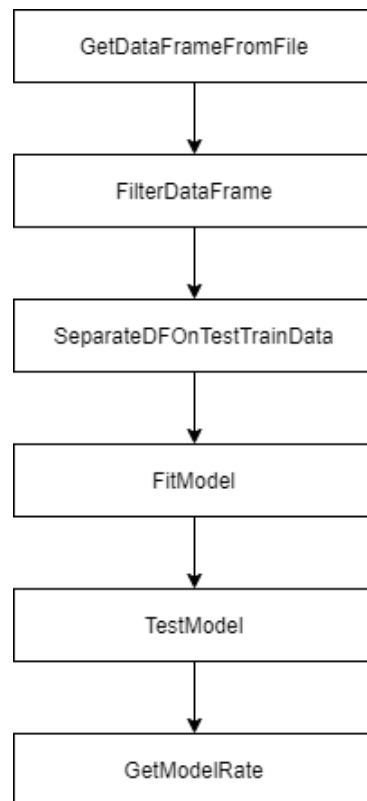


Рис 4.1 Послідовність виконання функцій реалізації запропонованого підходу

FilterDataFrame фільтрує данні: видалення пустих рядків даних, перевірка, щоб значення тестів кольорової преференції були від 1 до 12, еталонне значення від 1 до 4 , також виконується нормалізація даних, для методів LR, KNN, SVR.

SeparateDFOnTestTrain повертає структуру в якій знаходиться чотири масиви з розділеними даними для тренування і тестування моделі, також відокремлює вхідні та вихідні дані.

FitModel викликає метод тренування моделі, на основі методу опорних векторів з радіально базисною функцією ядра, приймає на вхід дані для тренування, на основі цих даних відбувається навчання моделі. Після цього необхідно викликати TestModel, яка приймає дані для тестування моделі, та повертає оцінку стану людини.

GetModelRate метод для оцінки якості роботи моделі - приймає на вхід отримані значення з попереднього кроку, та вірну оцінку стану з SeparateDFOnTestTrain та повертає коефіцієнт дисперсії, середньоквадратичну помилку, середню абсолютну помилку моделі навчену на основі методу машинного навчання.

4.3 Практичне застосування розробленого ПЗ

Для реалізації ПЗ було використано мову програмування Python, оскільки має гарну продуктивність при обробці великих даних, має великий набір бібліотек, а також легко інтегруються з іншими мовами програмування. Так для Python існує реалізація інтерпретатора для JVM з можливістю компіляції, CLR, інші незалежні реалізації. Оскільки програмне забезпечення НАНЦ України розроблено на основі технології Microsoft .Net 4.7 та фреймворк WPF, то інтеграція розробленого ПЗ можливе за допомогою виклику .Net драйвера який запускає Python скрипти.

Було використано бібліотеку Scikit-learn, для використання реалізації різних методів машинного навчання, оскільки дана бібліотека має широкий вибір алгоритмів навчання з учителем, а також має детальну документацію. Також була використана бібліотека Pandas для групування даних. Дана

бібліотека перш за все призначена для очищення і первинної оцінки даних за загальними показниками, також містить набори даних типів `DataFrame`, які застосовуються в якості вхідних типів в модулі машинного навчання.

Програмна реалізація підходу до аналізу психофізіологічного стану людини на основі кольорової преференції та суб'єктивних відчуттів виконано на основі процедурного підходу, використання підходу ООП являється недоцільним, оскільки об'єм коду невеликий. Такий підхід дозволить гнучко використовувати розроблене програмне забезпечення. Наприклад для практичного застосування, функція розділення даних на тестові дані та дані для тренування непотрібна, оскільки при практичному застосування для навчання моделі буде використано весь набір даних, функція фільтрації даних може бути замінена, або взагалі її можна не використовувати, що призведе до зміни ефективності роботи моделі.

При практичному використанні розробленого програмного забезпечення мінімальною послідовністю для прогнозування стану людини буде послідовний виклик наступних функцій : `GetDataFrame`, оскільки необхідно отримати дані на основі яких буде проводитись навчання моделі, `FilterDataFrame`, необхідно прибрати дані, які будуть не валідні, оскільки це може призвести до некоректної роботи моделі, за необхідності дана функція може бути замінена, `FitModel`, безпосереднє навчання моделі і `TestModel`, повертає результат прогнозу стану людини на основі нових даних пройденого тесту кольорової преференції та еталонного результату кольорової преференції.

Результатом роботи розробленого програмного забезпечення є модель навчена на основі методу опорних векторів, яка прогнозує стан людини по наданим даним проходження тесту кольорової преференції та еталонного результату кольорової преференції , а також можливо отримати параметри

оцінки моделі: коефіцієнт дисперсії, середньоквадратичну помилку, середню абсолютну помилку.

Розроблене ПЗ було запущено у середовищі Jupyter Notebook у режимі розробника для підбору параметрів для методів машинного навчання, порівняння роботи методів машинного навчання для прогнозування стану людини (описаного в розділі 3) та перевірки працездатності.

Висновки:

1. Запропоновано програмну реалізацію розробленого підходу до аналізу стану людини на основі кольорової преференції та суб'єктивних відчуттів.
2. Програмну реалізацію запропонованого підходу до аналізу стану людини на основі кольорової преференції та суб'єктивних відчуттів виконано в вигляді модулю, що дозволить легко підключати його до стороннього ПЗ.
3. Розроблене ПЗ виконано на основі процедурного підходу з застосуванням бібліотек Scikit-learn та Pandas. Побудовано та описано функціональну діаграму розробленого ПЗ.

ЗАГАЛЬНІ ВИСНОВКИ ПО РОБОТІ

1. Проведено огляд дослідження психофізіологічного стану людини на основі кольорової преференції та аналіз сучасних технологій обробки даних.
2. Описано проблеми дослідження та коригування психофізіологічного стану людини та визначено необхідність у розробці підходу до аналізу стану людини.
3. Запропоновано підхід до аналізу психофізіологічного стану людини на основі кольорової преференції та суб'єктивного відчуття людини, який дозволить завчасно виявляти та реагувати на зміну психофізіологічного стану людини.
4. Досліджено ефективність методів машинного навчання для прогнозування стану людини. Обрано метод опорних векторів, оскільки при його дослідженні були отримані найкращі параметри оцінки для прогнозування стану людини.
5. Розроблено програмну реалізацію запропонованого підходу до аналізу стану людини на основі кольорової преференції та суб'єктивних відчуттів. Програмну реалізацію виконано в вигляді модулю, для можливості подальшої інтеграції в стороннє програмне забезпечення.
6. Прийнято участь у студентській конференції «ПРІТС-2020».

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Инновационные методы исследований в психофизиологии цветового восприятия: Методическое пособие / С.- А. И. Мадяр, Е. В. Моисеенко Е. Э. Ковалевская / – К., 2015. – с. К.
2. Мадяр С.-А.Й., Моисеєнко Є. В., Пишнов Г. Ю. і співавт. /Методика поліхромно-адаптаційної біорегуляції психофізіологічного стану людини (методичні рекомендації) / , МОЗ України, АМН України, Український центр наукової медичної інформації і патентно-ліцензійної роботи: Київ-2006.-32 с.
3. Мадяр А.Й. Спосіб корекції психофізіологічного стану людини. / А. Й. Мадяр, О.Е. Ковалевська, В.В. Арбатов, В.Н. Бержанський, М.В. Луцюк, Є. В. Моисеєнко, В.Б. Павленко, С.В. Чорний // Патент на корисну модель № 28058 від 26 листопада 2007 року.
4. Mitchell, T. Machine Learning. McGraw Hill, 1997.
5. Ron Kohavi; Foster Provost. Glossary of terms. Machine Learning 1998
6. Шитиков В. К., Мاستицкий С. Э. Классификация, регрессия, алгоритмы Data Mining с использованием R. ,2017